Using Logistic Regression Markov Chain-Based Machine Learning Models To Predict Basketball Using NBA Data

Thiyagarajan, Sherwin (School: Albuquerque School of Excellence)

My project is about modeling professional basketball with data science and machine learning. Data science and analytics currently have significant impacts on the NBA's best teams. NBA teams such as the Houston Rockets have used analytics to adopt a unique offensive style. With more research, machine learning and data science could impact major coaching decisions in professional basketball. For this project, I used a Logistic Regression Markov Chain model, or LRMC, as the base for a multi-layered model intended to predict team success. Initially developed for collegiate basketball, this model uses logistic regression to find elements used to calculate transition probabilities for a Markov chain. The resulting steady-state probabilities can then be used to rank teams in a league or predict the outcome of head-to-head matchups. I formulated design criteria to make the model accurate, practical, accessible, and efficient. I created an LRMC model that met all of these criteria. Using data analysis tools within Python, I was able to make a valuable tool for predicting the outcome of games. From there, I added Monte Carlo simulations to validate the model and simulate across a broader range of teams and more complex scenarios such as playoff series.

Awards Won: